Examination of the Secondary Structure of Proteins by Deconvolved FTIR Spectra

D. MICHAEL BYLER and HEINO SUSI, *Eastern Regional Research Center, U.S. Department of Agriculture, Philadelphia, Pennsylvania 19118*

Synopsis

Fourier transform ir (FTIR) spectra of 21 globular proteins have been obtained at 2 cm^{-1} resolution from 1600 to 1700 cm^{-1} in deuterium oxide solution. Fourier self-deconvolution was applied to all spectra, revealing that the amide I band of each protein except casein consists of six to nine components. The components are observed at 11 well-defined frequencies, although all proteins do not exhibit components at every characteristic frequency. The root mean square (RMS) deviation of 124 individual values from the 11 average characteristic frequencies is 1.9 cm^{-1}. The observed components are assigned to helical segments, extended *beta*-segments, unordered segments, and turns. Segments with similar structures do not necessarily exhibit band components with identical frequencies. For instance, the lower frequency *beta*-structure band can vary within a range of approximately 15 cm^{-1}. The relative areas of the individual components of the deconvolved spectra were determined by a Gauss–Newton, iterative curve-fitting procedure that assumed Gaussian band envelopes for the deconvolved components. The measured areas were used to estimate the percentage of helix and *beta*-structure for each of 21 globular proteins. The results are in good general agreement with values derived from x-ray data by Levitt and Greer. The RMS deviation between 22 values (*alpha*- and *beta*-content of 11 *beta*-rich proteins measured by both techniques) is 2.5 percentage points; the maximum absolute deviation is 4 percentage points.

INTRODUCTION

Ir spectroscopy constitutes one of the earliest experimental methods for estimating the secondary structure of polypeptides and proteins.[1] Because water absorbs strongly in the most important spectral region, around 1640 cm^{-1}, studies in aqueous solution are difficult unless deuterium oxide is used as a solvent.[2] Even then only qualitative information is usually obtained, because components of absorption bands associated with specific substructures such as the helical regions and *beta*-strands cannot be resolved.[3] Some semiquantitative results have, nevertheless, been obtained by judicious band-fitting with the help of digital computers.[3] Application of Fourier self-deconvolution to ir[4,5] and Raman[6] spectra has removed many of the previous limitations, and has made more detailed studies of globular proteins possible.[7,8] This development has been facilitated by the application of second-derivative spectroscopy.[9,10]

The purpose of the present study is to investigate to what extent deconvolved ir spectra in the 1600- to 1700-cm^{-1} region (the amide I region) can be used to investigate the secondary structure of globular proteins in aqueous solution in a quantitative manner. Although crystallographic structure studies of proteins have made great progress over the past several years, experimental structure studies of globular proteins in their native aqueous environment are still in a relatively primitive state. It appears that a judicious application of deconvolved ir spectra in the amide I region can form a useful supplement to the existing methods, particularly because many proteins of great biological interest have not yet been crystallized.

BASIC CONSIDERATIONS

The Secondary Structure of Globular Proteins

Proteins are frequently referred to as having a certain fraction of helical structure and a certain fraction of extended *beta*-structure, but there are no generally valid objective criteria to establish numerical values in this context because, among other things, the endpoints of segments with different secondary structures are not easy to determine in an objective manner. The best we can do is establish well-defined criteria and then apply them in a consistent fashion.[11—13] On the basis of Cartesian atomic coordinates, obtained by x-ray crystallography, several such sets have been established and algorithms have been written to evaluate the secondary structure in a consistent fashion. A comprehensive approach has been given by Levitt and Greer,[11] who use dihedral angles and bond distances associated with the *alpha*-carbon atoms, in conjunction with hydrogen-bonding considerations, to describe the secondary structure of 60 proteins in some detail. Kabsch and Sander use primarily hydrogen-bonding criteria to describe the secondary structure of 62 proteins.[12] Amato and Liebman have proposed a promising method based on "distance matrix analysis"[13] but have not published results covering a sufficiently large number of proteins to permit a thorough evaluation. To provide a comparison of results obtained by different approaches, Table I presents data on the *beta*-strand content and the helix content of the protein papain, as obtained by different methods. This protein was chosen as an example because it contains appreciable amounts of both helical segments and *beta*-strands, and it has been investigated by several different mathematical and experimental procedures.

The data in Table I show that the results for papain vary from method to method, particularly for the proportion of *beta*-structure. The overall Fourier transform ir (FTIR) results for this protein agree best with the results of Levitt and Greer.[11] Several reasons may account for the differences that do occur. One is the difficulty in for-

TABLE I
Helix Content and *Beta*-Structure Content of Papain as Determined by Different Methods

	X-ray[a]	X-ray[b]	X-ray[c]	CD[d]	FTIR[e]
Total beta-structure	.29	.16	.14	.05	.32
Total helix	.29	.19	.28	.27	.27

[a] Levitt and Greer (Ref. 11).
[b] Kabsch and Sander (Ref. 12).
[c] Original estimate by Drenth et al. (Ref. 14).
[d] Value from CD by Provencher and Glockner (Ref. 15).
[e] This investigation.

mulating precise but general definitions for each type of conformation. Another probably arises from the empirical basis sets adopted as standards in some types of calculations. For example, synthetic homopolypeptides may not serve as good structural and spectroscopic models for actual proteins.

Of the two methods based on the interpretation of x-ray results, the one by Kabsch and Sander[12] is more elaborate and detailed, while the one by Levitt and Greer[11] is easier to grasp. As seen below, the latter is also in good agreement with our spectroscopic results. While truly objective criteria for describing the secondary structure of globular proteins are difficult to establish, clearly defined rules can nevertheless be formulated. In this investigation we adhere to the system of Levitt and Greer[11] as far as helical segments and *beta*-strands are concerned. Bends and turns are defined somewhat differently by various authors.[11-16] As given by Levitt and Greer[11] these substructures are not easily associated with specific ir bands, although such bands are clearly observed in deconvolved as well as in second-derivative spectra.[7,8,10]

Spectroscopic Consideration

In the ir spectra of proteins, the secondary structure is most clearly reflected by the amide I and amide II bands, particularly the former, which absorbs around 1620 to 1690 cm^{-1} and is primarily associated with the stretching vibrations of peptide carbonyl groups.[1-3] There are basically two theoretical approaches available to evaluate the existing data:

(a) The first is a perturbation treatment based on Wilson's *GF* matrix formulation, as applied to a weakly coupled oscillator model, developed by Miyazawa[17] and refined by Krimm and co-workers.[18,19] The treatment is based on fundamental concepts developed for polymer spectra[20] and is related to the factor-group approach for molecular crystals.[21] The basic requirements are (1) that the spectroscopic repeat unit must be small compared to the wavelength of the radiation and (2) that the system must be large enough to approximate an infinite

array of identical subunits.[20,21] The first criterion is satisfied for ir spectra of proteins, but the second is usually not for most of the helices and *beta*-strands of globular proteins. The thoroughly studied lysozyme molecule provides a good example. The *beta*-structure of this protein can be represented by six segments with 3, 3, 6, 6, 4, and 3 residues, respectively.[11] Each segment is separated from the next by residues with a different secondary structure. The only more or less regular "sheet" consists of only three short strands,[16] as seen in Fig. 1. Similarly, the twenty *beta*-strands of carbonic anhydrase consist of 3 to 11 residues each.[11] Ten of the twenty *beta*-segments form a twisted sheet whose overall size is less than 10 × 10 residues (Ref. 16, p. 295). The remaining segments are isolated and form no clearly discernible sheets. There are also a total of nine helical segments, ranging in length from 5 to 21 residues[11] (representing about 1.5 to 6 turns of a helix). The perturbation treatment, which assumes segments of infinite length, can thus provide only a rough approximation for the spectra of actual globular proteins.

Another problem is presented by the concepts of "parallel chains"

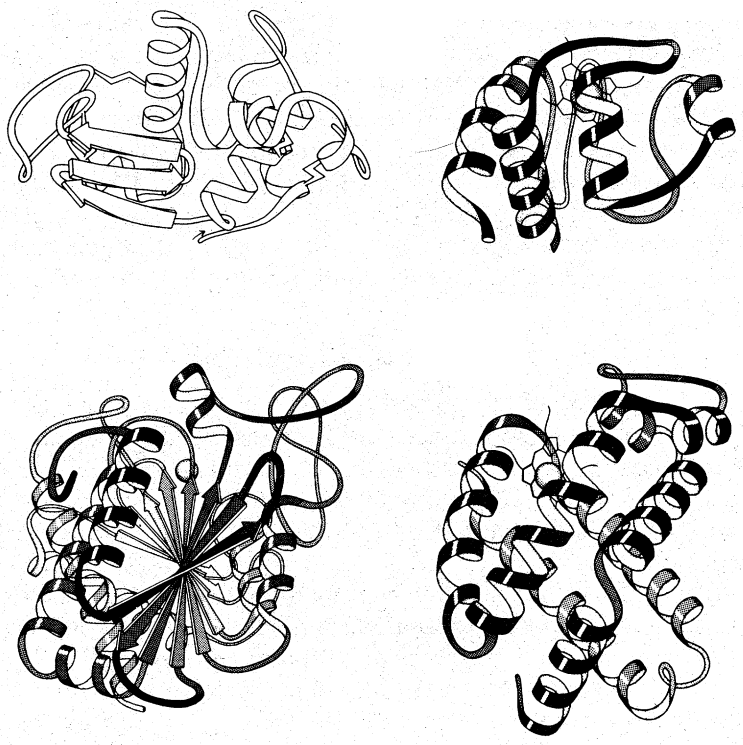

Fig. 1. Some schematic examples of the folding of globular proteins. Upper left: egg lysozyme. Upper right: cytochrome *c*. Lower left: carboxypeptidase. Lower right: hemoglobin *beta*-subunit. (From Jane S. Richardson, Ref. 15. Copyright by Academic Press.)

and "antiparallel chain pleated sheets."[17] In carbonic anhydrase, the relative directions of the ten segments forming the central sheet can be given as: $--+-+--+++-++$ (Ref. 16, p. 294). Carboxypeptidase (Fig. 1) has a similar twisted sheet with the form[11,16]: +10, −10, +13, +10, +6, +10, −7, −8. Here + and − also refer to the relative direction of adjacent strands. In neither case is it possible to talk about well-defined parallel or antiparallel pleated sheets as assumed by the perturbation treatment.[16-18] (The third *beta*-strand in carboxypeptidase, for instance, has one neighboring chain running in a parallel direction, while the other adjacent strand is aligned antiparallel.)

(b) The second approach consists of a straightforward normal coordinate analysis of polypeptide models representing the *alpha*-helix,[22,23] the *beta*-structure,[23] and "turns."[24,25] In globular proteins, different segments of a given structural class within the same molecule can exhibit absorptions at different frequencies. For instance, in the spectrum of concanavalin A the lower frequency *beta*-structure band has components at 1622, 1634, and 1639 cm^{-1}. In proteins where this band is not split, the band center varies from 1633 cm^{-1} in the spectrum of ribonuclease S to 1638 cm^{-1} in the spectrum of trypsin. (See text below and Table II). While detailed normal coordinate calculations are of great value in understanding the spectra, they cannot generally predict exactly the observed frequencies of globular proteins, because the structure of the latter is less regular than that found for polypeptide models. We base our assignments on previous experimental work[2,3] as well as on calcuated values.[23-25] In the latter case we must allow some leeway for deviations caused by the irregular nature of the secondary structure of globular proteins as compared to synthetic polypeptides.

Deconvolution and Band Fitting

The algorithm we use for deconvolution is based on the one described by Kauppinen et al.[4] It assumes a Lorentzian shape for the original, unresolved components[4] and it is most easily applicable if all bands have the same width. In actual spectra of proteins neither assumption is necessarily satisfied and application of the deconvolution procedure becomes an empirical process. For quantitative studies, deconvolution must be held to a minimum if side lobes are to be avoided and the constancy of band areas preserved.[4,7] We found values of 13 cm^{-1} for the full bandwidth at half height (FWHH) and 2.4 for the resolution enhancement factor[4] to be satisfactory, if the signal-to-noise ratio is better than 500. Similar values were found to yield good results by Griffiths and co-workers in an independent study at the University of California.[7] In exceptional cases, for molecules with particularly narrow or broad bands, we use supplementary values of deconvolution parameters for comparison.

TABLE II

Deconvolved Amide I Frequencies and Assignments (cm^{-1})

Protein	Extended Chains				Helix	Unordered		Turns and Bends			
	Low Components			High Components							
Carbonic anhydrase	1636		1625	1678	1653	1645	1660	1668		1690	
Carboxypeptidase	1635	1628	1623	1679	1652	1642	1665	1671		1690	
Casein						1644		1671			
α-Chymotrypsin	1637		1627	1674	1654	1647	1665		1681	1688	
Chymotrypsinogen	1637		1627	1674	1654	1646	1664		1682	1688	1696
Concanavalin A	1639	1634	1623	1671	1653	1646	1659		1684		1694
Elastase	1639	1633	1620	1673	1657	1645	1665		1683		
Immunoglobulin G	1637		1624	1672	1651		1659			1688	
α-Lactalbumin	1637		1627	1676	1652	1644	1665		1684		
β-Lactoglobulin		1634	1623	1679	1654	1646	1665		1685		1692
Lysozyme		1630		1673	1654	1641	1666		1682		
Papain	1640	1632	1621	1679	1654	1646	1662	1670		1687	
Ribonuclease A	1637	1628		1676	1655	1646	1665		1684	1688	
Ribonuclease S		1633		1672	1653	1645	1663		1681		
Trypsin	1638		1627	1673	1654	1646	1664		1681		
Trypsinogen	1636		1624	1675	1654	1646	1663		1684		1693
Trypsin inhibitor (soybean)	1636			1675	1656	1646	1664		1685		
Rounded Average	1637	1631	1624	1675	1654	1645	1663	1670	1683	1688	1694
RMS Deviation	1.4	2.5	2.4	2.6	1.5	1.6	2.2	1.4	1.5	1.1	1.7
Maximum Absolute Deviation	3	3	4	4	3	4	4	2	2	2	2

Frequency positions of the band centers can be independently evaluated by second derivative spectroscopy[10] but these are not necessarily identical with the centers of unresolved bands, although they are in most cases very close if the band separation is not too small compared to the bandwidth.

The deconvolution procedure, which includes both an apodization function and a line-shape function, alters the shape of the resulting resolution-enhanced bands.[4] We find, empirically, that the deconvolved spectra can be fit reasonably well by assuming that the components have Gaussian line shapes. Another problem is presented by the unknown integrated intensities (the area absorptivities) of the different component bands. It is well known that they are not equal.[17-19] For instance, the antiparallel chain pleated sheet results in one strong ir active band (ca. 1637 cm^{-1}), one weak band (ca. 1675 cm^{-1}), and one very weak band. Our study of over 20 proteins (11 of them also investigated by Levitt and Greer[11]) suggests, on an empirical basis, that the sum of all the integrated areas of the "*beta*-bands" as a fraction of the total amide I band area is closely related to the total "*beta*-content" of a given protein. The same applies to helices. Otherwise, quantitative estimations would be very difficult or even impossible, because the number and frequency of the components varies from protein to protein.

Although several approximations are involved, we do not depend on any transferred secondary-structure values, model polypeptides, or sets of basis functions and statistical correlations. Once the band assignments have been made, each determination is an individual event.

EXPERIMENTAL PROCEDURE

Many of the protein samples were obtained from the Sigma Chemical Company: Carbonic anhydrase, bovine erythrocyte, C-7500; carboxypeptidase A, bovine pancreas, C-0386; *alpha*-chymotrypsin, bovine pancreas, C-7762; *alpha*-chymotrypsinogen A, bovine pancreas, C-4879; concanavalin A, jack bean (*Canavalia ensiformis*), C-2010; cytochrome *c*, equine heart, C-7752; elastase, porcine pancreas, E-0127; ferritin, equine spleen, F-4503; hemoglobin, bovine erythrocyte, H-2500; lysozyme, chicken egg white, L-6876; myoglobin, sperm whale skeletal muscle, M-0380; papain, papaya latex, P-4762; ribonuclease A, bovine pancreas, R-5500; ribonuclease S, bovine pancreas, R-6000; trypsin, porcine pancreas, T-0134; trypsinogen, bovine pancreas, T-1143; trypsin inhibitor, soybean, T-9003. *Alpha*-casein, *alpha*-lactalbumin, and *beta*-lactoglobulin A were obtained through the courtesy of Harold M. Farrell, Jr., of our research center. These last three proteins were all isolated from bovine milk.

Sample preparation and instrumentation have been previously described.[7,8,10] To summarize: Samples were generally prepared as so-

lutions in 0.01*M* $NaCl/D_2O$ with a protein concentration of approximately 5% (w/v). (Carboxypeptidase was an exception and was dissolved in 2.4M $LiCl/D_2O$.) They were allowed to equilibrate for at least 24 h or until the amide II band around 1550 cm^{-1} had disappeared[2] before running the spectra used for this study. Deuteration shifts the amide II band by about 100 cm^{-1}, but displaces the amide I band by a much smaller amount (typically less than 5-10 cm^{-1}). Our work is based consistently on deuterated proteins. All spectra of these protein solutions were obtained from 4000 co-added interferogams run on a Nicolet 7199 FTIR spectrometer. Nominal instrument resolution was set at 2 cm^{-1}; the cell path length was 0.075 mm.

Second-derivative spectra were obtained for all proteins as reported earlier (Fig. 2).[10] The frequencies of the band centers thus measured were used as initial input parameters for the band-fitting procedures described below. Deconvolution was carried out by assuming an initial

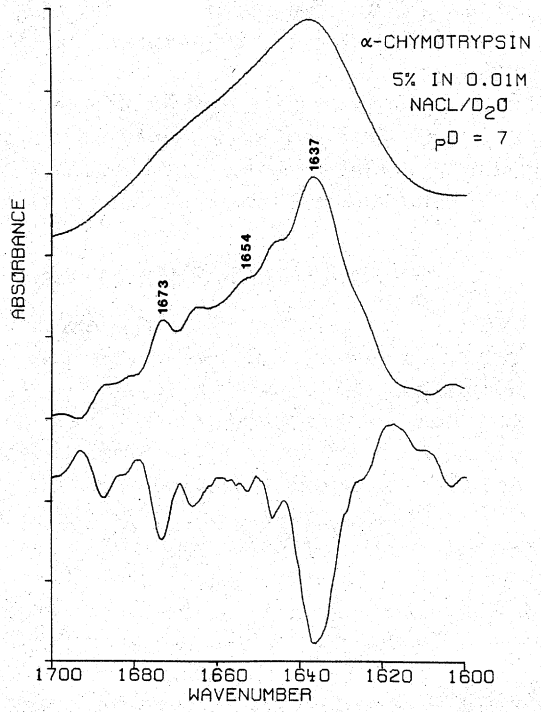

Fig. 2. Original FTIR spectrum, deconvolved spectrum, and second-derivative spectrum of *alpha*-chymotrypsin, 1600-1700 cm^{-1}. 5% w/v in C_2O with 0.01*M* NaCl. pD 7. Path length 0.075 mm. Deconvolution parameters VFO = 13 cm^{-1}, VF1 = 2.4 (see text).

Lorentzian line-shape function with a FWHH of 13.0 cm^{-1} (defined in the Nicolet software as VF0) and a resolution enhancement factor[4] of 2.4 (VF1 in the Nicolet program). For some samples, values of 10.0 cm^{-1} and 2.0, or 18.0 cm^{-1} and 2.8, respectively, were also used for comparison. Both parameters usually must be changed simultaneously to optimize this resolution-enhancement procedure, and to minimize spectral artifacts and distortion. If the values used for VF0 and VF1 are too high, band areas may become distorted.[4] We therefore deliberately chose not to risk overdeconvolution and limited ourselves to the use of conservative values of VF0 = 13.0 cm^{-1} and VF1 = 2.4.

To measure the relative areas of the partially resolved amide I band components, the deconvolved spectra were curve fit. This was accomplished by means of ABCUS, a FORTRAN program written and developed by William C. Damert and his colleagues at the Eastern Regional Research Center's computer center, which makes use of a Gauss–Newton iterative procedure. Gaussian band shapes were assumed for the deconvolved components, as previously discussed. When there are a large number of variables, the program will not necessarily provide a unique solution. In most cases "false" solutions involve an unreasonably large value for the width of one band, as well as a rather high Root mean square (RMS) error, and can be easily rejected.

The procedure was in general carried out as follows: (1) Only components detected by second derivative spectra were first considered. Frequencies were fixed at values determined by second derivatives. Half widths at half height (HWHH) were assumed to be around 4 cm^{-1} and were also fixed. Peak intensities were adjusted manually with the help of the monitor screen. (2) All intensities were iterated to obtain a minimum RMS error. (3) All intensities and frequencies were fixed at the obtained values, all widths were iterated. (4) All frequencies were iterated and other variables kept constant. (5) If an obvious gap remained, an additional component was added (usually around 1630 cm^{-1} for the *beta*-structure or around 1646 for "unordered" segments). The cycle was repeated, if necessary, until a satisfactory fit was obtained. (It is generally not possible to start by simultaneously iterating all parameters because the calculation may diverge.)

The areas of some components are easier to evaluate than others. The spectrum of ribonuclease S (Fig. 3) furnishes an example. The value of the area of the 1633-cm^{-1} band is well defined by the shape of the deconvolved spectrum. There is more ambiguity in the areas of the 1653- and 1646-cm^{-1} bands, although the very low RMS error observed (0.6% of the maximum absorbance) strongly supports the given solution. In general, the areas of the components around 1635 cm^{-1} (associated with the *beta*-structure) are more easily measured than the components close to 1653 or 1646 cm^{-1}, associated with helical portions and "random" portions, respectively.

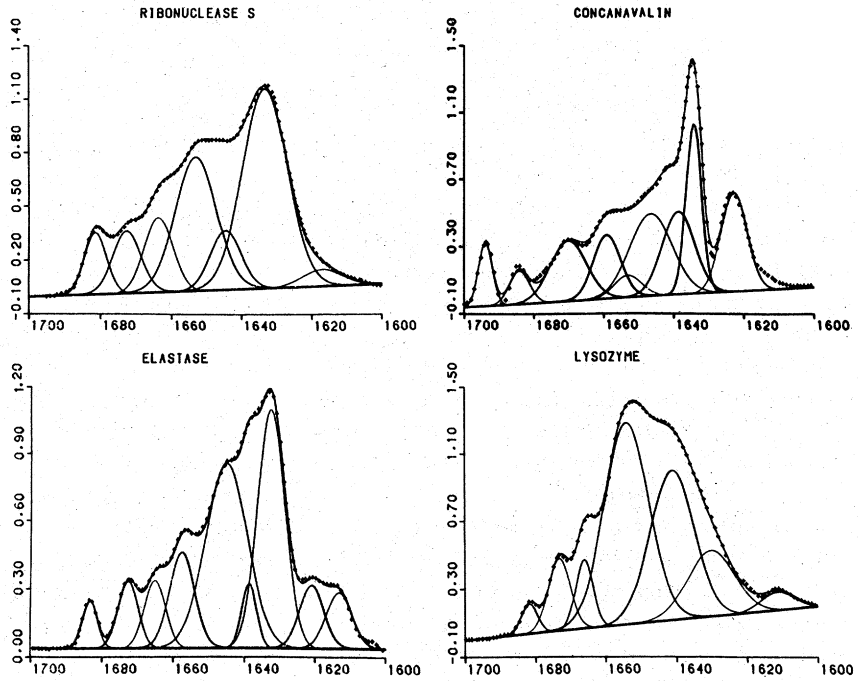

Fig. 3. Deconvolved amide I bands of four globular proteins with well-known secondary structures containing *alpha*- and *beta*-segments. [Absorbance plotted vs wavenumber (cm^{-1}).] Experimental conditions as in Fig. 2. Curve fitting carried out assuming Gaussian band shapes for deconvolved components. (Before deconvolution, no fine structure of the bands is detectable and no meaningful band fitting is possible.)

RESULTS AND DISCUSSION—ASSIGNMENTS

General Assignment of Deconvolved Amide I Components

Figure 2 gives the original spectrum, the deconvolved spectrum, and the second-derivative spectrum of chymotrypsin in the amide I region to provide an indication of how the two resolution-enhancement procedures complement each other. Figure 3 shows the deconvolved spectra of four proteins that contain both *alpha*-helices and *beta*-segments, and whose structures have been analyzed by Levitt and Greer.[11] Figure 4 displays the deconvolved spectra of three proteins with a very high fraction of *alpha*-helix, as well as the spectrum of the "orderless"[3] protein *alpha*-casein. Figure 5 displays the spectra of four more proteins of mixed conformation. To conserve space we reproduce the deconvolved spectra of only some of the proteins studied. Those selected have been chosen as typical.

Table II lists the frequencies of amide I components, after deconvolution, of 17 globular proteins that contain both *beta*-strands and

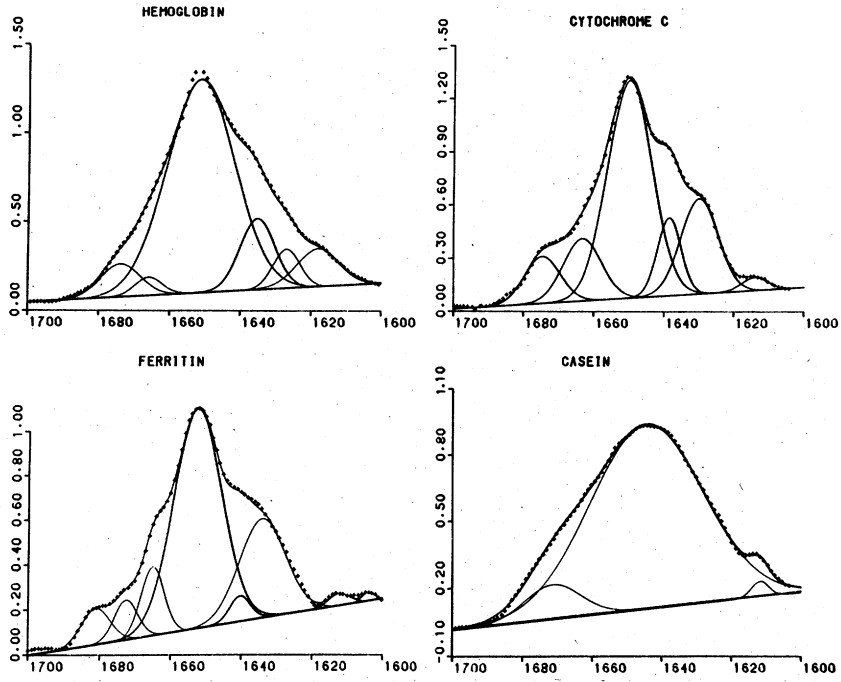

Fig. 4. Deconvolved amide I band of three proteins with very high helix content, and of the apparently unordered protein casein. [Absorbance plotted vs wavenumber (cm^{-1}).] Experimental conditions and curve fitting as in Fig. 3.

helical segments. We begin the discussion with the amide I components associated with *beta*-strands.

The "single-chain approach" predicts two ir active amide I components for *beta*-segments.[26] They can be roughly described as the "out-of-phase" and "in-phase" stretching vibrations of two neighboring peptide C=O groups. (These vibrations correspond to those classified as A_1 and B_1 symmetry species for the C_{2v} factor groups to polyglycine.) A more elaborate model that includes interchain hydrogen-bonding interactions predicts three ir active modes, one of them very weak.[17-19] In globular proteins the *beta*-strands are usually quite short,[11] and the sheets that exist often involve only a limited number of chains, sometimes as few as two or three.[11,16] The more complex approximations[17-19] therefore do not necessarily apply in this case. Several low-frequency (out-of-phase) *beta*-components, however, are frequently observed between 1620–1640 cm^{-1} in the spectrum of a single protein. By contrast, generally a single high-frequency (in-phase) component is observed in the region of 1670–1680 cm^{-1}. (Our decision to assign this band to the *beta*-strands rather than to turns is based on the spectra of proteins with very high *beta*-contents like concanavalin A and immunoglobulin G.)

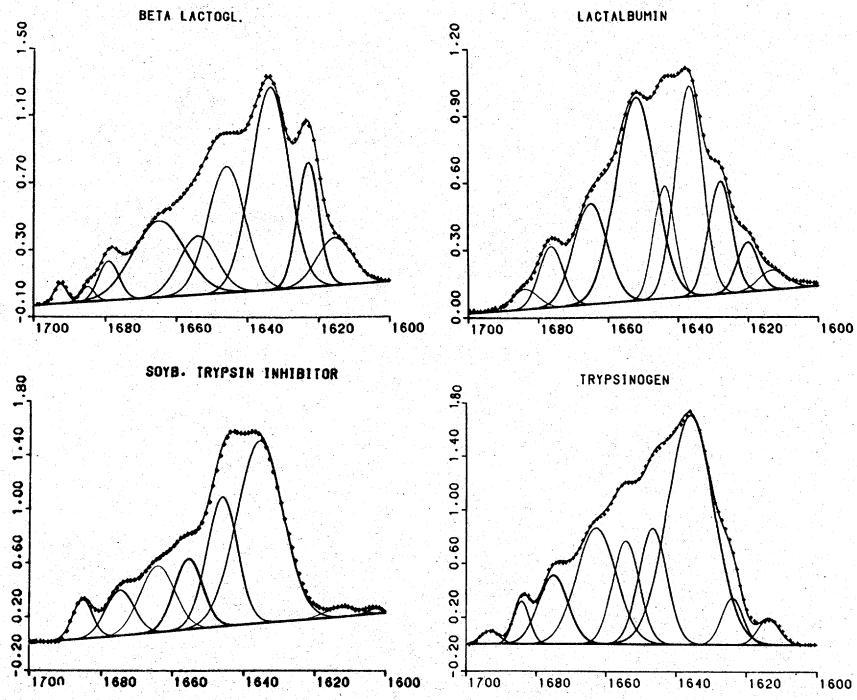

Fig. 5. Deconvolved amide I bands of four globular proteins with partially known secondary structures. [Absorbance plotted vs wavenumber (cm^{-1}).] Conditions as in Figs. 3 and 4.

The first four columns in Table II list these "*beta*-frequencies." Some proteins (e.g., concanavalin A, elastase, papain) have as many as three low-frequency *beta*-components, ranging from 1620 to 1640 cm^{-1}. This assignment is strongly supported by band-area measurements, as described below. It is tempting to try to associate the different low-frequency branches with particular classes of segments or substructures, but this does not appear to be easy. The two varieties of ribonuclease furnish a good example. The A-form exhibits bands at 1637 and 1628 cm^{-1}, while the S-form has a single one at 1633 cm^{-1}, despite the considerable similarity between the secondary structures of the two forms.[11,16] (It is possible that the rather broad 1633-cm^{-1} component of the S-form actually consists of two subcomponents that are not resolved by our techniques.)

Most proteins in Table II exhibit a component around 1654 cm^{-1}, associated with helical segments.[1-3,23] Several spectral components are theoretically predicted for the *alpha*-helix.[23] Three are actually observed by vibrational dichroism spectra of *alpha*-helical polypeptides.[27] Only one band is observed near this frequency in either the deconvolved or the second derivative spectra of the proteins studied in this investigation.

A component observed in most proteins close to 1645 cm^{-1} is as-

signed to "unordered segments" by comparison with the band center of the apparently "orderless" *alpha*-casein and with alkali-denatured *beta*-lactoglobulin[8] (see Fig. 4 and Table II). This component probably results from peptide groups that are hydrogen bonded to solvent molecules but not to other groups within the protein.[2,3] Most proteins in Table II also exhibit a band close to 1663 cm^{-1}, which we assign to turns, along with bands around 1670, 1683, 1688, and 1694 cm^{-1}, in general agreement with theoretical calculations for peptides.[24,25]

Deconvolution of the amide I band of globular proteins thus produces distinct components at 11 wavenumber values; the observed frequencies for each characteristic absorption exhibit RMS deviations from the mean of 1.1 to 2.6 cm^{-1}. (All proteins do not, of course, display bands at all 11 characteristic frequencies.) The proposed assignments are supported by the relative values of integrated intensities as determined from the curve-fit data and discussed below.

Amide I Components of Highly Helical Proteins

Table III lists the amide I components of four highly helical proteins after deconvolution. In addition to the helix-band near 1650 cm^{-1}, all four proteins show bands around 1627–1638 and 1671–1675 cm^{-1}. Because neither hemoglobin nor myoglobin has *beta*-structure in the generally defined sense,[11,16] these bands must be associated with the short, extended chains connecting the helical cylinders (e.g., the residues 79–84, 98–99, 120–123, and 150–153 in myoglobin[11]). These segments consist of 2 to 6 residues each[11], they are not bent into "turns"[28,29] nor can they form "sheets" of any kind. They are not classified as *beta*-strands by any reported method of protein structure analysis. If pairs of neighboring C=O oscillators within such chain segments interact strongly, the observation of amide I (C=O stretching) bands near 1635 and 1675 cm^{-1} would, nevertheless, be in accord with the generalized single-chain approximation used to describe the spectra of well-defined *beta*-strands. Similar reasoning would also apply to the spectra of cytochrome *c* and ferritin (see below.)

In addition to the strong band close to 1650 cm^{-1} and the weak pair around 1635 and 1675 cm^{-1}, helical proteins show a weak band close to 1664 cm^{-1}, which is assigned to turns like the corresponding bands

TABLE III
Amide I Frequencies for Highly Helical Proteins (cm^{-1})

Protein	Helix	Extended Chains				Turns	
		Low Component		High Component			
Hemoglobin	1651	1634	1627	1675		1665	
Myoglobin	1650		1630	1672			
Cytochrome *c*	1650	1638	1630	1675		1664	
Ferritin	1653	1634		1673		1665	1681

in Table II. This band is very weak in hemoglobin and was not observed at all in myoglobin. Ferritin has an additional band at 1681 cm^{-1}, also probably associated with turns (cf. Table II).

QUANTITATIVE EXAMINATION OF SECONDARY STRUCTURE

Beta, *Alpha+Beta*, and *Alpha/Beta* Proteins Analyzed by Levitt and Greer

Table IV gives the calculated percentage of helix and *beta*-strands for 11 of the proteins covered by Levitt and Greer[11] (hereafter abbreviated as L & G). The agreement between L & G values and those calculated from FTIR data is quite good for the *beta*-content as well as the helix content, as seen in columns 1 to 3 and 5 to 8 of Table IV. The agreement of our FTIR values with original x-ray estimations and with results obtained by CD is not as good. It is somewhat suprising that estimates of secondary structure based on precisely defined interpretations of Cartesian atomic coordinates[11] agree so well with data obtained by the purely experimental FTIR method, which does not depend on any transferred reference standards. For turns, comparison of the FTIR estimates with those from other procedures is more difficult. No well-characterized numerical experimental values appear to exist except for a few scattered CD data.[15]

The percentages of *beta*-structure were calculated by adding the areas of all bands assigned to *beta*-segments and expressing the sum as a fraction of the total amide I band area. The summation of the 1635-cm^{-1} region components and the 1675-cm^{-1} component leads to good results for proteins where all of the *beta*-segments are antiparallel, as well as for proteins where some strands lie parallel. In the main *beta*-sheet of carbonic anhydrase several chains are parallel, and in the case of carboxypeptidase most of the adjacent chains in the central *beta*-sheet are parallel $(+-++++--)$,[11,16] as seen in Fig. 1. Yet these proteins exhibit *beta*-bands in the same frequency ranges as proteins with only antiparallel chains. This empirical observation conflicts with the most recently published theoretical calculations.[19] To our knowledge no unambiguous experimental data concerning the spectroscopic properties of parallel *beta*-sheets have previously appeared in the literature. We are presently engaged in experimental studies that should help resolve this problem.

A quite similar procedure was applied to estimating helical segments and to turns. Such an approach obviously neglects possible differences between the integrated intensities (area absorptivities) of amide I branches arising from segments with different secondary structures. It does, nevertheless, lead to quite acceptable empirical results for the proteins studied.

TABLE IV

Quantitative Estimation of the Secondary Structure of Some *beta*, *alpha+beta* and *alpha/beta* Proteins Covered by Levitt and Greer[a]

Protein	% *beta*-structure				% Helix			
	FTIR	X-ray[b]	Dif[c]	CD[d]	FTIR	X-ray[b]	Dif[c]	CD[d]
Carbonic anhydrase	49	45	+4	—	13	16	−3	—
		40		—		17		—
Carboxypeptidase	33	30	+3	15	40	39	+1	43
		15		0		37		45
α-chymotrypsin	51	49	+2	29	12	10	+2	9
		34		53		9		5
Chymotrypsinogen	49	46	+3	—	13	11	+2	—
		—		36		—		9
Concanavalin A	60	60	0	41	4	2	+2	8
		51		46		2		25
Elastase	45	47	−2	49	11	10	+1	4
		52		46		—		0
Lysozyme	21	19	+2	21	41	45	−4	45
		16		29		41		32
Papain	32	29	+3	5	27	29	−2	27
		14		0		28		29
Ribonuclease A	50	46	+4	44	21	22	−1	26
		40		39		23		21
Ribonuclease S	50	53	−3	37	25	23	+2	25
		44		33		26		24
Trypsinogen	54	56[e]	−2	—	11	9[e]	+2	—

[a] Ref. 11.

[b] The first x-ray value presented for each protein is that calculated by Levitt and Greer (Ref. 11); the second value is that of the original investigator. (See Ref. 11 for citations).

[c] Dif = the difference between the FTIR value and the x-ray value. The RMS of the deviations between FTIR values (this work) and x-ray values (Levitt and Greer, Ref. 11) is 2.5 percentage points.

[d] The first CD value for each protein is from Ref. 15; the second is from Ref. 30, except for chymotrypsinogen (Ref. 37).

[e] X-ray values for trypsinogen are for the very similar protein trypsin-DIP as given by Ref. 11.

The average absolute difference between values obtained by L & G[11] and by the deconvolved FTIR method is 2.3 percentage points for all calculated values of helix content and *beta*-content. The RMS difference is 2.5 percentage points. Table I likewise shows that a comparison of the FTIR results with those obtained by any other method, either based on crystallographic data or on spectroscopic measurements, leads to considerably larger differences than a comparison with the results of L & G.

Proteins with High Helix Content

These proteins are treated separately because of some previously discussed spectral features. (See the section above on *amide I components of highly helical proteins*.) Table V provides information regarding four proteins with a helix content of about 50% or higher. In contrast to the proteins described in Table IV, the helix content obtained for heme proteins by FTIR is lower than indicated by L & G, but quite close to the estimation of the original x-ray investigators. It appears that the residues at the ends of helical segments that are included by L & G (but not by the original investigators) do not contribute to the spectroscopically observed helix band centering close to 1650 cm^{-1}. The relatively high bandwidth of the main component of the heme proteins (about 18 cm^{-1} FWHH) further suggests that the helical sections are nonuniform and that the broad band probably consists of several subcomponents that remain unresolved even after deconvolution.

The 1635- and 1675-cm^{-1} region bands observed in these proteins apparently arise from the entire nonhelical portions of the molecules, except some well-defined turns. Cytochrome *c* provides a good example.

TABLE V
Estimation of the Secondary Structure of Some Proteins with High Helix Content

Protein	% Helix			% Extended Chain	% *beta*-structure	
	FTIR	X-ray[a]	CD[b]	FTIR	X-ray[a]	CD[b]
Cytochrome *c*	51	49	37	34	10	9
		39	44		—	0
Ferritin	57	—	—	29	—	—
		74[c]	52[d]		11[c]	37[d]
Hemoglobin	74	86	—	25	0	—
		75	68[e]		—	—
Myoglobin	76	87	86	24	0	—
		77	80		—	—

[a] Sources of x-ray values as in Table IV, unless otherwise specified.
[b] Sources of CD values as in Table IV, unless otherwise specified.
[c] Data from Ref. 35.
[d] Data from Ref. 38.
[e] Data from Ref. 36.

The helix content is in reasonable agreement with L & G. Figure 1 indicates three well-defined turns,[16] corresponding to 12 residues and consisting of 11.7% of the entire chain of 103 residues. The area of the 1664 cm^{-1} band indicates about 12% turns. The remaining parts of the molecule are represented by bands at 1630, 1638, and 1674 cm^{-1}, which can be associated with the less well-defined extended-chain segments depicted clearly in Fig. 1.

Proteins not Covered by Levitt and Greer

Table VI provides data for five more proteins with a low helix content. These proteins have not been studied by L & G[11] and, in some cases, have not been studied at all by high-resolution x-ray crystallography. A rough comparison is provided with whatever data are available from various sources. Casein constitutes a special case because the native form appears not to fold into any discernible, regular pattern of secondary structure.[2,3,31] Its deconvolved amide I spectrum, as shown in Fig. 5, as well as its second derivative spectrum, exhibit a broad band centering at 1644 cm^{-1}. The extreme broadness of this band (32 cm^{-1} FWHH) again suggests the presence of a number of closely spaced unresolved components, with the strongest component centering at 1644 cm^{-1}, close to the unordered component of other proteins (cf. Table II).

Quantitative Secondary-Structure Examinations—Summary

FTIR spectroscopy, utilizing deconvolved amide I bands in conjunction with band fitting by Gaussian components, yields results that are in good agreement with the interpretations of x-ray data carried out

TABLE VI
Estimation of the Secondary Structure of Some Proteins not Included in Previous Systematic Studies

Protein	% *beta*-structure			% Helix		
	FTIR	X-ray	CD	FTIR	X-ray	CD
Casein	0	—	—	0	—	13[a]
Immunoglobulin G	76	70[b]	—	9	3[b]	—
α-lactalbumin	41	—	15[c]	33	—	26[c]
β-lactoglobulin	50	—	50[d]	10	—	15[d]
Trypsin	55	56[e]	36[d]	16	9[e]	—
Trypsin inhibitor (soybean)	52	(50)[f]	—	10	(0)[f]	—

[a] Ref. 31.
[b] Values for immunoglobulin G Fab' (Ref. 11) given for comparison.
[c] Ref. 32.
[d] Ref. 33.
[e] Values for trypsin-DIP (Ref. 11) given for comparison.
[f] Qualitative estimates from Ref. 34.

by Levitt and Greer[11] for proteins classified as *beta*, *alpha+beta*, and *alpha/beta*. The FTIR method reported here thus provides a means for rapidly estimating the secondary structures for proteins that have not been studied in any detail by x-ray crystallography or have not been crystallized at all. Very similar spectra and equally good results are obtained for proteins with *beta*-regions composed of antiparallel chains and for proteins, such as carboxypeptidase, where most *beta*-strands in the main sheet are parallel (Fig. 1).

In some samples the areas of components associated with major structural segments are easily determined, while some uncertainty remains about the areas of minor components. Ribonuclease *S* (Fig. 3) furnishes an example. The areas representing *beta*-components (bands centering at 1633 and 1672 cm^{-1}) are well defined, but uncertainty remains concerning the relative areas of the bands centering at 1645 cm^{-1} (unordered regions) and 1663 cm^{-1} (turns) because of insufficient resolution. Higher deconvolution would increase the apparent resolution, but could easily lead to distorted spectra and substantial errors in area measurements.

For heme proteins the helix content as determined by FTIR is lower than estimated by the methods of Levitt and Greer,[11] but is in good agreement with the original estimates from x-ray investigations. In the case of highly helical proteins that also contain ill-defined strands, the latter absorb like the *beta*-segments of the proteins covered in Table II. For proteins with a high helix content, the FTIR value for *beta*-structure content represents an upper limit, which might also include the ill-defined extended chains between the helical segments.

Dr. William Damert and Salvatore N. Amato of the Eastern Regional Research Center's computer center generously provided much invaluable advice and assistance. We are indebted to Dr. Hamish N. Munro, Director of the U.S.D.A. Human Nutrition Research Center on Aging at Tufts University, Boston, Massachusetts, for many helpful discussions and especially for calling to our attention the latest reports on the structure of ferritin. Janine N. Brouillette, Howard T. Fertman, and Stephen T. McGady carefully prepared many of the protein solutions and obtained their FTIR spectra. To each of these persons we express our sincere thanks.

References

1. Elliott, A. & Ambrose, E. J. (1950) *Nature* **165**, 921–922.
2. Timasheff, S. N., Susi, H. & Stevens, L. (1967) *J. Biol. Chem.* **242**, 5467–5473.
3. Rüegg, M.. Metzger, V. & Susi, H. (1975) *Biopolymers* **14**, 1465–1471.
4. Kauppinen, J. K., Moffatt, D. J. Mantsch, H. H. & Cameron, D. G. (1981) *Appl. Spectrosc.* **35**, 271–276.
5. Griffiths, P. R. (1983) *Science* **222**, 297–302.
6. Thomas, G. J., Jr. & Agard, D. A. (1984) *Biophys. J.* **46**, 753–768.
7. Yang, W.-J., Griffiths, P. R., Byler, D. M. & Susi, H. (1985) *Appl. Spectrosc.* **39**, 282–287.
8. Purcell, J. M. & Susi, H. (1984) *J. Biochem. Biophys. Methods* **9**, 193–199.
9. Kauppinen, J. K., Moffatt, D. J., Mantsch, H. H. & Cameron, D. G. (1981) *Anal. Chem.* **53**, 1454–1457.

10. Susi, H. & Byler, D. M. (1983) *Biochem. Biophys. Res. Comm.* **115**, 391–397.
11. Levitt, M. & Greer, J. (1977) *J. Mol. Biol.* **114**, 181–239.
12. Kabsch, W. & Sander, C. (1983) *Biopolymers* **22**, 2577–2637.
13. Amato, S. V. & Liebman, M. N. (1982) *Biophys. J.* **37**, 386a.
14. Drenth, J., Jansonins, J. N., Koekoek, R. & Wolthen, B. G. (1971) *Adv. Prot. Chem.* **25**, 79–115.
15. Provencher, S. W. & Glockner, J. (1981) *Biochemistry* **20**, 33–37.
16. Richardson, J. S. (1981) *Adv. Prot. Chem.* **34**, 167–339.
17. Miyazawa, T. (1960) *J. Chem. Phys.* **32**, 1647–1652.
18. Krimm, S. (1962) *J. Mol. Biol.* **4**, 528–540.
19. Krimm, S. & Abe, Y. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 2788–2792.
20. Krimm, S., Liang, C. Y. & Sutherland, G. B. B. M. (1956) *J. Chem. Phys.* **25**, 549–561.
21. Halford, R. S. (1946) *J. Chem. Phys.* **14**, 8–15.
22. Rabolt, J. F. Moore, W. H. & Krimm, S. (1977) *Macromolecules* **5**, 1065–1074.
23. Dwivedi, A. M. & Krimm, S. (1984) *J. Phys. Chem.* **88**, 620–627.
24. Krimm, S. & Bandekar, J. (1980) *Biopolymers* **19**, 1–29.
25. Bandekar, J. & Krimm, S. (1980) *Biopolymers* **19**, 31–36.
26. Fukushima, K., Ideguchi, Y., & Miyazawa, T. (1963) *Bull. Chem. Soc. Jpn.* **36**, 1301–1307.
27. Lipp, E. D. & Nafie, L. A., (1985) *Biopolymers* **24**, 813–823.
28. Venkatachalam, C. M. (1968) *Biopolymers* **6**, 1425–1436.
29. Crawford, J. L., Lipscomb, W. N. & Schellmann, C. G. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 538–542.
30. Chang, C. T., Wu, C.-S., & Yen, Y. T. (1978) *Biochemistry* **91**, 13–31.
31. Creamer, L. K., Richardson, T. & Parry, D. (1981) *Arch. Biophys. Biochem.* **211**, 689–691.
32. Robbins, F. M. & Holmes, L. G. (1970) *Biochim. Biophys. Acta* **221**, 234–240.
33. Creamer, L. K., Parry, D. A. D. & Malcolm, G. N. (1983) *Arch. Biophys. Biochem.* **227**, 98–105.
34. Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H. & Blow, D. M. (1974) *Biochemistry* **13**, 4212–4228.
35. Rice, D. W., Ford, G. C., White, J. L., Smith, J. M. A. & Harrison, P. M. (1984) in *Protides of the Biological Fluids, Colloquium* **31**, Peters, H., Ed., Pergamon Press, New York/London.
36. Hennessey, J. P., Jr. & Johnson, C. W. (1981) *Biochemistry* **20**, 1085–1094.
37. Chen, Y. H., Yang, J. T. & Martinez, H. M. (1972) *Biochemistry* **11**, 4120–4131.
38. Listowsky, I., Blauer, G., Englard, S. & Betheil, J. J. (1972) *Biochemistry* **11**, 2176–2181.
39. Sawyer, L., Papiz, M. Z., North, A. C. T. & Eliopoulos, E. E. (1985). *Biochem. Soc. Trans.* **13**, 265.